IMM-NYU 275
OCTOBER 1960

NEW YORK UNIVERSITY
INSTITUTE OF
MATHEMATICAL SCIENCES

448 900

On The Foundations of Statistical Inference. II

ALLAN BIRNBAUM

61-1-3
XEROX

PREPARED UNDER
CONTRACT NO. NONR-285(38)
WITH THE
OFFICE OF NAVAL RESEARCH
UNITED STATES NAVY

IMM-NYU 275
October 1960

New York University

Institute of Mathematical Sciences

ON THE FOUNDATIONS OF STATISTICAL INFERENCE. II

Allan Birnbaum

This report represents results obtained at the Institute of Mathematical Sciences, New York University, under the sponsorship of the Office of Naval Research, Contract No. Nonr-285(38).

1960

CONTENTS

0. Introduction and summary. Some principal technical developments of Part I of this paper are derived here in more elementary fashion, under the restriction to statistical experiments with discrete sample spaces, but under the more general condition that any finite number of (simple) statistical hypotheses may be represented. For any such experiment, it is shown that for typical purposes of informative statistical inference, just the likelihood function on an observed outcome can and should be reported and interpreted to provide inferences of general interest concerning the statistical hypotheses (or unknown parameter values); and that for such purposes, the structure of the experiment from which an outcome was obtained is irrelevant, apart from determination of the likelihood function. Specific techniques for interpretation of likelihood functions are developed, particularly "intrinsic confidence methods" which constitute an appropriate generalization and refinement of confidence methods and conditional confidence methods. The relations of such methods to traditional methods based on the " principle of insufficient reason", are discussed, as to form and interpretation. In Sections 9-11, analogous developments are given for experiments involving translation or scale parameters.

1. <u>The canonical form of an experiment</u>. We consider a given experiment E, assuming that questions of experimental design, including those of choice of a sample size or possibly a sequential sampling rule, have been dealt with, and that the sample space of possible outcomes x of E is a specified set $S = \{x\}$. We assume that each of the possible distributions of X is represented by a specified elementary probability function $f_i(x)$: if the hypothesis H_i is true, the probability that E yields an outcome x in A is

$$P_i(A) = \int_A f_i(x)d\mu(x) ,$$

where μ is a specified σ-finite measure on S, and A is any measurable set.

We assume here that a finite number k of hypotheses are under consideration: $H_1, \ldots H_k$, $k \geqq 2$. We shall omit comments on the particular features of the case of binary experiments, $k = 2$, which were discussed in Part I; and we shall refer to Part I at some points where methods or interpretations are immediately applicable without complication to experiments with $k > 2$.

For any binary experiment E, let

$$r_{ij} = r_{ij}(x) = \log[f_i(x)/f_j(x)], \quad i,j = 1, \ldots k, \ i \neq j.$$

Let

$$r = r(x) = [r_{12}(x), \ldots r_{1k}(x), r_{23}(x), \ldots r_{2k}(x), \ldots r_{(k-1),k}(x)].$$

It is well-known that r is a sufficient statistic, which may or may not be minimal sufficient, depending upon the structure of E. (If $f_i(x) = f_j(x) = 0$ or ∞, we define $r_{ij}(x) = 0$. The statistic r contains components which are redundant in many experiments; for example, if $0 < f_2(x) < \infty$ for all x, then for all x we have $r_{13}(x) = r_{12}(x) + r_{23}(x)$. However it is convenient to tolerate such possible redundancies for purposes of general discussion, and to take account of them appropriately for more specific purposes.)

Let

$$F_i(r) = \text{Prob}\ [r(X) \leqq r \mid H_i], \quad i = 1, \ldots k,$$

where the inequality between vectors denotes the corresponding inequality between respective coordinates: $r_{ij}(x) \leqq r_{ij}$. In general, $F_i(r)$ is a generalized multivariate distribution function, and $r(X)$ is a generalized multivariate random variable, in the sense that some coordinates of the latter may assume infinite values with positive probability under some hypotheses. The set of k distribution functions F_i of the statistic r may be taken as a canonical form of any experiment.

For some purposes, a different canonical representation of an experiment may be more convenient. For example, let $k = 3$, let $g = (g_1, g_2, g_3)$ be any set of <u>formal</u> prior probabilities for the respective hypotheses. Let $g_i^*(x)$ denote the formal posterior probability of H_i, given that $X = x$, for each i and x. Let $d(x)$ be any (measurable) function such that, for each x, the value of $d(x)$ is the index ($i = 1, 2$, or 3) of

the hypothesis H_i (or one of the hypotheses) with greatest posterior porbability, given $X = x$; that is, $d(x)$ is a formal Bayes solution, with respect to the prior distribution g. Let

$$\alpha_i = \text{Prob}\,[d(X) \neq i | H_i],\ i = 1,2,3 \text{ and let } \alpha = (\alpha_1, \alpha_2, \alpha_3).$$

Then α is the set of error-probabilities α_i, under respective hypotheses H_i, of the inference (or decision) function $d(x)$. A basic part of the theory of statistical decision functions is the investigation, for various experiments, of the set of all such points α, under all possible choices of g. It is well-known that, for each g, the set of such points is either a single point, or a line-segment, or a convex subset of a plane (that is, a convex set, of dimension at most $(k-1) = 2$); and that, for all g, the set is a convex surface in the unit cube. (The preceding assumes tacitly, without essential loss of generality, that x includes an observation on an auxiliary randomization variable.) But

$$1 - \alpha_1 = \text{Prob}[r_{12}(X) \mathbin{.\!\geq\!.} \log(g_2/g_1),\ r_{13}(X) \mathbin{.\!\geq\!.} \log(g_3/g_1) | H_1],$$

$$1 - \alpha_2 = \text{Prob}[r_{12}(X) \mathbin{.\!\leq\!.} \log(g_2/g_1),\ r_{23}(X) \mathbin{.\!\geq\!.} \log(g_3/g_2) | H_2], \text{ and}$$

$$1 - \alpha_3 = \text{Prob}[r_{13}(X) \mathbin{.\!\leq\!.} \log(g_3/g_1),\ r_{23}(X) \mathbin{.\!\leq\!.} \log(g_3/g_2) | H_3],$$

for any $d(x)$ corresponding to a given g; here the inequality symbols $.\!\geq\!.$ and $.\!\leq\!.$ refer to the arbitrary definition of $d(x)$ on points where equality holds. In experiments for which the distributions of components $r_{ij}(X)$ are continuous, these equations define a unique point $\alpha = \alpha(g)$ for each g; since $r_{13}(x) \equiv r_{12}(x) + r_{23}(x)$

the distribution $F_1(r)$ of $r(X)$ under H_1 is represented directly by the above equation for $1-\alpha$, when g is varied over its range. Similarly the distributions $F_2(r)$ and $F_3(r)$ are represented by the equations for $1-\alpha_2$ and $1-\alpha_3$. The same interpretations can be given, with attention to details, in cases of discontinuous distributions. Thus the convex surface of points α is a canonical form of an experiment, which is convenient for some purposes.

In the next sections we shall for the most part consider experiments with discrete distributions $f_i(x)$ (or, slightly more generally, discrete distributions $F_i(r)$). For our purposes, it will be convenient to represent each such experiment E by a stochastic matrix:

$$E = (p_{ij}), \quad i = 1,\ldots k, \quad j = 1,\ldots m; \sum_{j=1}^{m} p_{ij} = 1 \text{ for each } i.$$

Here $p_{ij} = \text{Prob}\,[X = j|H_i]$; the sample space is the range of j; m may be finite or infinite; and in the latter case the range of j can when convenient be taken to be the doubly-infinite sequence of integers, $-\infty < j < \infty$.

Redundancies in such representations of experiments may be eliminated as follows, when desired: (A) If two columns of such a matrix are proportional ($p_{ij} = cp_{ih}$ for some $j \neq h$ and some c, for each i), these columns may be deleted and replaced by the single column having elements $(p_{ij} + p_{ih})$, with an appropriate revision of the subscripts j. (Since the probability of X = j, given that X = j or h, is independent of i, the "simpler"

experiment is not less informative.) If all such possible simplifications are made, j is a minimal sufficient statistic. (B) Since a permutation of columns represents a re-labeling of sample points, experiments differing only in this respect are equivalent. (C) When convenient, a standard manner of ordering columns may be adopted.

2. Some algebra of statistical experiments. Except where the contrary is indicated, we assume that experiments for some fixed number k of hypotheses are under consideration. An experiment $E = (p_{ij})$ will be called simple if its matrix has (after the simplifying operations described above) at most k columns; that is, a simple experiment has not more sample points (after simplification) than hypotheses. The completely informative experiment is (equivalent to) the identity matrix of order k; the uninformative experiment is (equivalent to) the single-column matrix with elements all unity; all other experiments (not necessarily simple) are called incompletely informative. Any experiment having two identical rows ($p_{ij} = p_{hj}$ for some $i \neq h$ and all j) will be called degenerate; even many replications of such an experiment are without value for distinguishing between certain of the hypotheses. An experiment E is called at least as informative as an experiment E^*, or is said to contain E^*, if there exists a stochastic matrix $Q = (q_{ij})$ such that $E^* = EQ$; that is, if $p^*_{ij} = \sum_{h=1}^{m} p_{ih}\, q_{hj}$. It is known that E contains E^* if and only if the convex hypersurface of points α, which constitutes a canonical form of E in the sense illustrated in the preceding Section, encloses the convex hyper-

surface of points α^* corresponding to E^*. The relation "contains" determines a partial ordering of all experiments for k hypotheses.

Let E_h, h = 1, 2, ..., denote any sequence of experiments, and let $g = (g_1, g_2, \dots)$ be any sequence of probabilities, $\sum_h g_h = 1$, assigned formally to the respective experiments. Then $E = \sum_h \oplus g_h E_h$ represents "the mixture g of the experiments E_h": the experiment E consists of the observation of a value h of the random variable H with distribution g, followed by use of the corresponding experiment E_h. If each of the "component" experiments E_h can be represented by a finite matrix (p^h_{ij}), then E is easily represented (before possible simplification) by the matrix (consisting of successive finite blocks)

$$E = \left[(g_1\, p^1_{ij}), (g_2\, p^2_{ij}), \dots \right].$$

<u>Example 1.</u> Let $E^* = \begin{pmatrix} .6 & .2 & .2 \\ .2 & .6 & .2 \\ .2 & .2 & .6 \end{pmatrix}$ and let $E^{**} = \frac{1}{7}\begin{pmatrix} 1 & 3 & 3 \\ 3 & 1 & 3 \\ 3 & 3 & 1 \end{pmatrix}$.

Let $\bar{E}$ be the mixture g = (5/12, 7/12) of these respective experiments. Then

$$\bar{E} = \frac{5}{12} E^* \oplus \frac{7}{12} E^{**} = \frac{1}{12}\begin{bmatrix} 3 & 1 & 1 & 1 & 3 & 3 \\ 1 & 3 & 1 & 3 & 1 & 3 \\ 1 & 1 & 3 & 3 & 3 & 1 \end{bmatrix}.$$

It is readily verified that E has an alternative decomposition represented by $\bar{E} = \frac{1}{3} E_1 \oplus \frac{1}{3} E_2 \oplus \frac{1}{3} E_3$, where

$$E_1 = \frac{1}{4}\begin{bmatrix} 3 & 1 \\ 1 & 3 \\ 1 & 3 \end{bmatrix}, \quad E_2 = \frac{1}{4}\begin{bmatrix} 1 & 3 \\ 3 & 1 \\ 1 & 3 \end{bmatrix}, \quad E_3 = \frac{1}{4}\begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 3 & 1 \end{bmatrix}.$$

An experiment will be called cyclic-symmetric (abbreviated "c.s.") if it can be represented in the form $E = (p_{ij}) = (A_1, A_2, \ldots)$ where each A_h is a square cyclic-symmetric matrix. (A square matrix (a_{uv}) of order k is cyclic-symmetric if its elements satisfy $a_{uv} = a_{u+1,v+1}$ and $a_{kv} = a_{1,v+1}$, for $u,v = 1,\ldots(k-1)$.) Examples are the experiments E^*, E^{**}, and $\bar{E}$ of Example 1 above.

Lemma 1. Every experiment $E = (p_{ij})$ is a component of some cyclic-symmetric experiment.

Proof: Let $E_h = (p^h_{ij})$, where $p^h_{ij} = p_{i-1+h,j}$, for $h = 1,\ldots k$, where a subscript exceeding k is to be diminished by k; thus $E_1 = E$. Let

$$\bar{E} = \sum_{h=1}^{k} \oplus \frac{1}{k} E_h.$$

Even if E is not finite, it is possible to order the columns of $\bar{E}$ so as to exhibit its cyclic symmetry, thus: the first k columns of $\bar{E}$ are respectively the initial columns of $E_1, \ldots E_k$, each multiplied by $1/k$; the next k columns of $\bar{E}$ are respectively the second columns of these matrices, multiplied by $1/k$; etc.

Example: Let $E = E_1$ of the preceding Example 1. Then E_2, E_3, and $\bar{E}$ are as defined in that example.

Lemma 2: Every cyclic-symmetric experiment is equivalent to a mixture of cyclic-symmetric simple experiments.

Proof: If E is cyclic-symmetric, it can be given the form $E = (A_1, A_2, \ldots)$, where each $A_h = (a^h_{ij})$ is a square cyclic-symmetric matrix. Let $g_h = \sum_{i=1}^{k} a^h_{i1}$ and let $E_h = (1/g_h) A_h$, for each h. Then E_h is a cyclic-symmetric simple experiment, and E admits the decomposition $E = \sum_h \oplus g_h E_h = (A_1, A_2, \ldots)$.

Example: In Example 1, $\bar{E}$ is c.s. and has the simple c.s. components E^*, E^{**}.

3. The partial ordering of simple cyclic-symmetric experiments.

For any fixed number k of hypotheses, we consider in this Section simple c.s. experiments and their partial ordering defined as above. For such experiments, E contains E^* if and only if $E^* = EQ$ for some square stochastic matrix Q; our primary purpose is to interpret this partial ordering more explicitly in terms of the forms of the c.s. matrices representing such experiments. It will suffice for our purpose to illustrate the general case by a detailed discussion of the case $k = 3$.

Form 1: Let $E_L = c \begin{bmatrix} 1 & L & L \\ L & 1 & L \\ L & L & 1 \end{bmatrix}$, where $c = 1/(1+2L)$ and L is any number satisfying $0 \leqq L \leqq 1$. The product of two such matrices, with respective parameters L and L', is easily found to be $E_{L''}$, which is of the same form with $L'' = (L+L'+LL')/(1+2LL')$. We have $L'' = 1$ only if L or $L' = 1$. We have $L'' \geqq \max(L,L')$, with strict equality only if L or $L' = 0$. Thus the class of experiments of form 1 is simply ordered, with smaller values of L representing more informative experiments.

Form 2: Let $E^*_L = c^* \begin{bmatrix} L & 1 & 1 \\ 1 & L & 1 \\ 1 & 1 & L \end{bmatrix}$, where $c^* = 1/(2 + L)$ and L is any number satisfying $0 \leq L \leq 1$. The product of such a matrix with one of form 1 above is easily found to be an experiment $E^*_{L''}$ of form 2, with $L'' = (L + 2L')/(1 + L' + LL') \geqq L$. Thus the class of experiments of form 2 is simply ordered, with smaller values of L representing more informative experiments.

<u>Form 3.</u> Let $E = c \begin{bmatrix} 1 & L & L' \\ L' & 1 & L \\ L & L' & 1 \end{bmatrix}$, where $c = 1/(1+L+L')$, with $0 \leq L' \leq L \leq 1$. This includes the preceding forms as special cases and in fact includes all c.s. simple experiments with $k = 3$. We proceed to consider the partial ordering of such experiments, writing $E = (p_{ij})$.

For each such experiment E, let $\alpha = p_{12} + p_{13}$. It is easily verified that the parameter α of E is the common coordinate of one of the points of the $(\alpha_1,\alpha_2,\alpha_3)$-surface which represents E in the canonical form described in Section 1 above. In particular, α is the probability (under each hypothesis) of an error, when the rule of maximum likelihood is used to choose one hypothesis on the basis of one observation from E (with equiprobable randomization in cases of non-unique maxima of the likelihood function). This inference rule is an admissible one, since it is readily derived as a Bayes solution of the problem described, with respect to the uniform prior distribution $g = (1/3,1/3,1/3)$.

Again, for each such experiment E, let $\beta = p_{13}$. For the problem of giving a confidence set estimator which excludes at least one of the hypotheses and has maximum probabilities of including the true hypothesis, a Bayes solution with respect to the uniform prior distribution gives the maximum likelihood rule which excludes just the least likely hypothesis (with exclusion of one chosen by equiprobable randomization in cases of non-unique minima). The error-probability of this rule, under each hypothesis, is the parameter β of E.

A simple necessary condition for E to contain E^* is that $\alpha \leqq \alpha^*$ and $\beta \leqq \beta^*$; for if E failed to achieve error-probabilities at least as small as E^* in the two specific problems just described, it would fail to contain E^*. This condition may be described thus: if E contains E^*, then the distributions in E are at least as highly concentrated as those in E^*, in the sense that under each hypothesis the most probable outcome has probability at least as high, and the least probable outcome has probability at least as small, in E as in E^*.

To illustrate that the preceding condition is not sufficient for comparability of experiments, consider

$$E_1 = \begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{bmatrix}, \quad \text{and } E_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}.$$

The above condition is satisfied, since $\alpha_1 = \alpha_2 = \frac{1}{2}$ and $\beta_1 = \frac{1}{6} < \beta_2 = \frac{1}{4}$. Consider a third specific inference problem, that of giving a point-estimate of the parameter $i = 1, 2$, or 3, with estimators to be appraised in terms of their probabilities of each of the possible kinds of errors. (The parameter α is simply a total probability of incorrect estimates.) Any non-randomized estimator may be represented by a function $d = d(j)$ of the outcome j which takes values in the range of i. Any randomized estimator may be represented as a "mixture" of such non-randomized estimators; for example, if $d(j)$ and $d'(j)$ are non-randomized estimators, then $d''(j) = c\, d(j) \oplus (1-c)\, d(j)$ represents the randomized estimator which, when j is observed, takes the value $d(j)$ with probability c and takes the value $d'(j)$

with probability (1-c). For any estimator $d = d(j)$, possibly randomized, let $a_{iu} = a_{iu}(d) = a_{iu}[d(.)] = \text{Prob}[d(X)=u|H_i]$ if $u \neq i$, $a_{iu} = 0$ if $u = i$, for $u,i = 1,2,3$.
When E_2 is used, the (admissible maximum likelihood) estimator $d(j) = j$ has all error-probabilities $a_{iu} = \frac{1}{4}$, $u \neq i$. When E_1 is used, it can be verified that every estimator (including randomized estimators) has at least one error-probability exceeding $\frac{1}{4}$. Thus E_1 does not contain E_2.

For experiments of the general form 3, we offer here no conveniently-applicable necessary and sufficient conditions for comparability of experiments in terms of their parameters L and L'; nor will this be necessary for our purposes.

For $k > 3$, similar considerations are applicable. For example, among c.s. experiments of the form $p_{11} > p_{12} = p_{13} = \ldots = p_{1k}$ it is easily verified as above that there is a simple ordering by the parameter p_{11}; the latter parameter is the probability, under each hypotheses, that the most likely hypothesis will be the true hypothesis.

4. <u>Inference methods with intrinsic justifications</u>. In the preceding Section, for various c.s. simple experiments there were described a number of methods of statistical inference or decision-making, including point-and confidence-set estimators; in addition, a number of methods of testing hypotheses were represented implicitly by the confidence-set methods described, in virtue of well-known simple relations between the two kinds of methods. For each of these methods, a more or less complete description was given of the probabilities of the various

possible appropriate and inappropriate inferences or decisions under respective hypotheses. The complete description of such relevant probabilistic properties of a given inference method can in principle always be determined; and for a given purpose of application, various possible inference methods can in principle be evaluated and compared on the basis of such probabilistic properties. Such considerations are an extension of those discussed in detail for the case $k = 2$ in Sections 7-9 of the preceding Part I, B: "Inference methods with probabilistic justifications." Each such probability is defined directly in the experiment under consideration; and each such error- probability can be interpreted in terms of relative frequencies of errors, under respective hypotheses, in conceptually-possible indefinite repetitionsof the given experiment. We turn now to an extension of the preceding Part I, C: " Inference methods with intrinsic justifications". Since our discussion here takes a somewhat different form, it will complement the earlier discussion of the case $k = 2$; for many details of interpretation, reference to the earlier discussion may useful even in connection with cases $k > 2$.

Lemmas 1 and 2 of Section 2 above pay a basic role in support of the following interpretations. According to Lemma 2, any c.s. experiment E may be regarded as a mixture of c.s. simple experiments E_h. It follows that any outcome of E may be regarded as: (a) the selection of a component E_h of E, determined randomly according to probabilities g_h which are fixed, independently of the hypotheses; followed by: (b) the observation of a single outcome of the selected experiment E_h.

We observe: (1) The likelihood function on any observed outcome of E (that is, the column of the stochastic matrix E corresponding to any observed outcome of E) is necessarily the same as (proportional to) the likelihood function when that outcome is regarded as an outcome of a selected component E_h. (2) The likelihood function on any observed outcome of E determines, essentially uniquely, the form of the simple c.s. component E_h of E from which the observed outcome could have arisen. (A single column of a simple c.s. experiment, specified up to a constant of proportionality, determines the form of that experiment essentially uniquely.) (3) Since the selection of a particular component E_h of E provides no information relevant to the hypotheses (although it determines the strength and nature of relevant evidence which can be provided by an outcome of E_h), it follows that <u>for purposes of informative inference, any outcome of any c.s. experiment can and should be interpreted in the same way as if it were an outcome of the essentially unique simple c.s. experiment determined by the observed likelihood function</u>. The variety of possible and possibly-useful interpretations of outcomes of simple c.s. experiments was illustrated in part in the preceding Section 3; such interpretations were expressed there in terms of error-probabilities, admitting frequency interpretations, defined in the simple c.s. experiment under consideration.

To establish a similar conclusion for experiments which are not necessarily c.s., we use Lemma 1, and the notation of its proof, in Section 2 above. The evidential interpretation of any outcome $X = j$ of any experiment E_1 should clearly coincide with the evidential interpretation of the following

outcome: the random selection of E_1 as a component of any mixture experiment $\bar{E}$ (of which E_1 is in fact a component), followed by use of E_1 and observation of its outcome $X = j$. The introduction into our discussion of any such experiment $\bar{E}$ containing E_1 as a component is an arbitrary step; however, the preceding comment shows that this step does not affect the recognizable evidential status of any outcome j of E_1; and the following comments show that this step is useful in throwing additional light on the evidential character of such an outcome. We take $\bar{E}$ to have the form defined in the proof of Lemma 1: $\bar{E} = \sum_{h=1}^{k} \oplus \frac{1}{k} E_h$, where the components E_h are defined as before. We are considering the evidential character of outcome j of E_1, and we have agreed that this is the same as the evidential character of the outcome "E_1 and its j^{th} outcome" of the mixture experiment $\bar{E}$. But $\bar{E}$ is c.s., and therefore the conclusion established above is applicable to its outcome "E_1 and its j^{th} outcome". Thus we conclude: <u>For purposes of informative inference, any outcome of any experiment can and should be interpreted in the same way as an outcome of a simple c.s. experiment having the same likelihood function; the structure of the original experiment is irrelevant, apart from determination of the likelihood function on the observed outcome</u>.

If $k = 2$, a simple c.s. experiment is a symmetric simple binary experiment $E = \begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{11} \end{bmatrix}$, which may be characterized by the single value $\alpha = \beta = p_{12} \leq \frac{1}{2}$ of its "error-probabilities",

as discussed in Part I above. For $k = 3$, a simple c.s. experiment $E = (p_{ij})$, with $p_{11} \geqq p_{12} \geqq p_{13}$, may be considered characterized by the values of its "error-probabilities" p_{12}, p_{13}; for this case, evidential interpretations of various specific forms, and their qualitative and quantitative properties in relation to error-probabilities, were described in detail in Section 2 above. Similar considerations hold for $k > 3$.

5. <u>Intrinsic confidence methods.</u> One useful method of expressing part of the evidential meaning of an outcome of a simple c.s. experiment is by use of inference statements of the <u>confidence set</u> form. For any such experiment $E = (p_{ij})$, with $p_{11} > p_{12} > \ldots > p_{1k}$, the maximum likelihood estimator of the unknown hypothesis H_i is formally a confidence set estimator with confidence coefficient p_{11}. The two most likely hypotheses, on any observed outcome, constitute a confidence set with coefficient $(p_{11} + p_{12})$; and so on. The set which, on any outcome, includes all but the least likely hypothesis is a confidence set with coefficient $(1 - p_{1k})$. If, for example, $p_{11} = p_{12} > p_{13} > \ldots > p_{1k}$, then all such confidence sets except the first can be defined in the same way; construction of a maximum-likelihood confidence set consisting of a single hypothesis could also be given formally in this case, but would be of little interest for typical purposes of informative inference.

Such <u>maximum likelihood confidence sets</u> are optimum in the sense (a) that each such set-estimator has, under each hypothesis, the largest possible probability of including the true hypothesis, among all (possibly randomized) set-estimators

whose confidence-sets are restricted to contain the same number or fewer points (hypotheses); and in the sense (b) that each such set-estimator has probabilities of including various false hypotheses, when respective hypotheses are true, which cannot be strictly improved, except by reduction of the confidence coefficient.

Such confidence sets were illustrated for the case $k = 3$ in Section 2 above. The set of confidence coefficients of such estimators characterizes the structure of a simple c.s. experiment; for example, for $k = 3$, the respective confidence coefficients are p_{11} and $p_{11} + p_{12}$; from these values, we can immediately calculate p_{12} and p_{13}, and thus determine the form of $E = (p_{ij})$.

If the experiment E whose outcome is to be interpreted happens to be of the simple c.s. form, then inference methods of the preceding kinds are confidence methods (confidence set estimation methods) of the kind introduced by Neyman: the confidence coefficients and error-probabilities referred to are then defined directly in terms of the structure of $E=(p_{ij})$ as just described. These confidence coefficients and error-probabilities admit the usual frequency interpretations, in terms of conceptually possible repetitions of the given experiment E.

If the experiment E happens to be c.s. but not simple, then it is (by Lemma 2 of Section 2 above) equivalent to a mixture of simple c.s. experiments. In this case the conclusion of the preceding Section can be given the interpretation: Any outcome of E should be interpreted as an outcome of the

corresponding simple c.s. component of E; in other words, any outcome of E should be interpreted "conditionally" with the selected simple c.s. component of E as the experimental frame of reference. In such a case, when confidence methods like those described above are used, based upon the simple c.s. experiment determined by the likelihood function on an observed outcome of E, these methods are formally an example of conditional confidence methods. Conditional applications of inference methods of standard kinds are widely used, and are generally considered appropriate for purposes of informative inference, when an appropriate conditional experimental frame of reference is recognized. Decomposition theorems such as Lemma 2 and its analogues may be considered mathematical analyses of the structures of statistical experiments which extend considerably the range of recognizably appropriate conditional frames of reference for purposes of informative inference. The confidence coefficients and error-probabilities of such conditional confidence methods admit the usual frequency interpretations, as conditional probabilities, in terms of conceptually possible repetitions of the given experiment E, conditional on the selection of the particular simple c.s. component of E which corresponds to the observed outcome.

If the experiment E is not c.s., the conclusion of the preceding Section nevertheless supports interpretations of an outcome of E as if it were an outcome of the simple c.s. experiment E' determined by the likelihood function on the observed outcome. In general, E' is not a component of E; and if such interpretations are expressed, for example, by maximum likelihood confidence

methods based on E', then the confidence coefficients and error-probabilities of such methods, which are defined in E', will not in general be interpretable as probabilities or conditional probabilities defined in E. For this reason, we designate such methods in general as <u>intrinsic confidence methods</u>. Intrinsic confidence methods constitute an extension and generalization of confidence methods and conditional confidence methods, appropriate for purposes of informative inference. (An intrinsic confidence method can always be regarded as a conditional confidence method in the <u>hypothetical</u> formal sense that, in some hypothetical c.s. experiment which contains the given experiment E as a component (Lemma 1), the intrinsic confidence method is also a conditional confidence method. This comment should not be confused with the development of the principal conclusion of Section 4 above.)

It should be noted that for a given k, different outcomes may give the same intrinsic confidence set with the same intrinsic confidence coefficient, although these outcomes have different likelihood functions which do not coincide completely. In such a case, other intrinsic confidence sets based on the respective outcomes will fail to coincide, reflecting differences in likelihood functions. This illustrates that in general any single intrinsic confidence statement expresses only part of the evidential significance of an outcome, and is only an incomplete summary of the likelihood function.

6. An interpretation of the "principle of insufficient reason" .

In the method of treating statistical inference problems which was initiated by Bayes and Laplace, " uniform prior probabilities"were postulated for the respective statistical hypotheses under consideration, and the formal "posterior probabilities", calculated by Bayes' formula, were interpreted as giving inferences from observational data to the hypotheses in the absence of, or independent of, background knowledge or prior opinions concerning the hypotheses. Evidently the intention of those who initiated and have used this method has been to treat, in suitably objective and meaningful terms, the problem of informative inference, that is, the problem of evidential interpretation of experimental outcomes, as it occurs in empirical research situations. Following Laplace, the method was widely accepted during the nineteenth century. Analysis and criticism of the possible ambiguity of the notion of "uniformity" of prior probabilities, and of the unclear nature of such "prior probabilities" in general, has led to a general rejection of this method throughout the present century. (The use of prior probabilities, not in general " uniform", to express background knowledge and/or prior opinion, continues to be recommended by a distinguished minority of modern statisticians. However,such recommendations are not addressed directly to the problem of informative inference as described above; but to problems of using experimental outcomes, along with background knowledge, prior opinion, and information about specific

features of an inference situation such as goals, practical consequences, etc., in order to reach appropriate decisions or conclusions.)

It is at least a striking coincidence that inference methods based upon the "principle of insufficient reason", in problems having suitable symmetry (or analogous) properties, coincide in form (although they differ in interpretation) with modern inference methods derived without use of prior probability notions. For example, if an experiment E happens to be simple c.s., then formal assignment of prior probabilities, each equal to 1/k, to the hypotheses H_i, leads to "posterior most probable" sets of hypotheses which coincide with the (optimum) maximum likelihood confidence sets found above; and each such set has a posterior probability which is numerically equal to the corresponding confidence coefficient.

Now the analysis of preceding sections shows that for purposes of informative inference, whatever the structure of E, its outcomes can and should be interpreted as outcomes of corresponding simple c.s. experiments. When such an appropriate experimental frame of reference for interpreting an experimental outcome is adopted, as it can and should be to serve the apparent intention of those who initiated use of the "principle of insufficient reason", then the formal application of the latter principle can be regarded as a formal algorithm for calculating the intrinsic confidence sets and coefficients which themselves have the independent justifications given above; and the term "posterior probability (determined with use of uniform prior probabilities)" may be regarded as a traditional terminology, in

place of which we use the term " intrinsic confidence coefficient", with the meaning established in the preceding Section.

Thus the "principle of insufficient reason", in such problems and uses, must be regarded as one of those "principles" which, in various mathematical disciplines, have been recognized and used to obtain "correct" results, in advance of perfectly clear formulations of the problems considered and of the precise nature of "correct solutions" to such problems. (In experiments for an infinite number of hypotheses, the "principle of insufficient reason" has been interpreted and used, despite the technical difficulty of specifying the mathematical meaning of "uniform prior probabilities", in such a way that the likelihood function is taken to be the elementary "posterior probability function" with respect to some "natural", "uniform" measure on the parameter space; while it is not necessary for mathematical reasons that the latter be probability measures, there remains the question of interpretation and possible ambiguity of "natural" or "uniform". We defer discussion of experiments for an infinite number of hypotheses.)

In retrospect, the early broad usage of the term "probability" is seen to have embraced at least the following kinds of meanings which now seem clear and distinct (although we have frequent occasion to use several of them in discussion of a single problem of a single problem of inference):

(a) Probability as used above to specify mathematical models of statistical hypotheses; admitting conceptual frequency interpretations.

(b) Prior probability (not in general uniform; and related posterior probability), as sometimes used to express prior opinion and background knowledge, brought to an inference problem. This aspect of inference situations has not been represented in our discussion, because it is not a part of the problem of informative inference as such.

(c) Posterior probability (calculated from formal uniform prior probabilities; "principle of insufficient reason "). In place of this traditional usage, we have the preferable term "intrinsic confidence (coefficient)" which is defined as above in terms of the more basic likelihood function and the interpretations established for the latter. In brief, "intrinsic confidence", and the more basic "likelihood" with its interpretations, explicate and replace this traditional usage of "posterior probability". Similarly, "uniform prior probability" could well be replaced by "uniform prior likelihood", the latter denoting the constant likelihood function which properly represents absence of informative observations (or presence of hypothetical uninformative outcomes) at the outset of an experiment. Then the traditional usage would be represented intact, with the term "probability" replaced by "likelihood" throughout, except for usage (a), and with all attention directed as above to the usual likelihood function.

7. <u>An interpretation of Fisher's "fiducial argument"</u>.

For any experiment $E = (p_{ij})$, we can assume without loss of generality that the range of j is doubly-infinite, $-\infty < j < \infty$, so that each row of (p_{ij}) is doubly-infinite, with $\sum_{-\infty < j < \infty} p_{ij} = 1$ for each i. The range of i may be, until otherwise specified,

either finite ($1 \leqq i \leqq k$), or countably-infinite ($1 \leqq i \leqq \infty$), or doubly-infinite ($-\infty < i < \infty$). We note that j is a sufficient statistic (not in general minimal sufficient), as is any one-to-one function of j. Any real-valued function $t = t(j,i)$ of i and j is called a quasistatistic; a quasistatistic becomes a statistic when its argument i is given any fixed value. A sufficient quasistatistic is one which becomes a sufficient statistic when i is fixed, in turn, at each of its possible values. (A minimal sufficient quasistatistic is defined analogously.) A stationary quasistatistic is one which determines statistics each having the same distribution in the sense that, letting $H(t,i) = \text{Prob}\,(t(X,i) \leqq t \mid H_i)$, we have $H(t,i) \equiv H(t,1)$, for each t and i. A pivotal quasistatistic is one which is both stationary and sufficient.

If E is such that $p_{i-j} \equiv p_{i-1,j-1}$, i is called a translation parameter, since for each i the distribution of X under H_i coincides with the distribution of $(X+i-1)$ under H_1. We observe that in any such experiment, the quasistatistic $t(j,i) = j-i+1$ is pivotal.

In the case of a simple c.s. experiment for a finite number k of simple hypotheses, represented by a square matrix $E = (p_{ij})$, the quasistatistic

$$t(j,i) = \begin{cases} j-i+1, & \text{if the latter is positive,} \\ j-i+1+k, & \text{otherwise,} \end{cases}$$

is pivotal. If the k outcomes j of such an experiment are regarded as cyclically ordered, as k uniformly-spaced points on

the circumference of a circle, then i may be regarded as a translation parameter in an extended use of that term. We shall call the parameter i of any simple c.s. experiment a <u>rotation</u> parameter, since for each i the distribution of X (mod. k) under H_i coincides with the distribution of (X+i-1) (mod. k) under H_1.

If E is simple c.s., then each of its columns (possible likelihood functions) is, in the formal mathematical sense, a probability distribution over the possible values i of the unknown parameter. The same is true of any column of any experiment in which i is a translation parameter having a doubly-infinite range. If E is an experiment of one of these two forms, then an example or analogue of Fisher's "fiducial argument" gives the following definition: When an outcome X = j has been observed, take the j^{th} column of E as the "fiducial probability distribution" of the parameter i. (We note that, in the case of such experiments, these distributions coincide formally with those obtained by formal application of the "principle of insufficient reason" discussed in Section 6 above.)

The "fiducial argument" by which a "fiducial distribution" has usually been defined may be illustrated in the present case, of translation and rotation parameters i, with use of the pivotal quasistatistics (usually called, in this context, "pivotal quantities") t(j,i) defined above, as follows: For each i and integer t, we have Prob $(t(X,i) \leq t|H_i) \equiv H(t,i) \equiv H(t,1)$, which is independent of i. H(t,1) is formally a cumulative probability distribution with argument t; if X = j is an observed outcome of E, then the function G(i,j) of i defined by G(i,j) = 1 - H(t(j,i)-, 1) is formally a cumulative probability distribution function, termed

the "fiducial distribution of the parameter i, when X = j has been observed". In the present cases we have the corresponding discrete elementary "fiducial probability function" $g(i,j) \equiv G(i,j) - G(i-1,j) = p_{ij}$, as stated above, which coincides with the likelihood function. In such cases, probability statements about i based formally on the fiducial probability function g(i,j) must parallel in form both confidence statements and statements based formally on the "principle of insufficient reason".

Fiducial methods were developed by Fisher evidently for the purpose of treating what we have called the problem of informative inference. For experiments for a finite number of hypotheses, our conclusion in Section 5 above shows that an appropriate frame of reference for informative inferences is always provided by a simple c.s. experiment determined by the likelihood function on the observed outcome. Evidently the adoption of such a frame of reference would serve the general intention for which fiducial methods have been developed. Adoption of such a frame of reference would also extend considerably the scope of formal applicability of the "fiducial argument", since the conditions of its applicability are evidently not met in many experiments (for k hypotheses), but in any simple c.s. experiment fiducial probabilities can be defined formally, as above; such adoption would lead to fiducial probability statements about i which always parallel in form intrinsic confidence statements defined in Section 5 above.

The preceding discussion of fiducial methods has been restricted to formal definitions, and to an opinion concerning the general purpose to which the methods are addressed. It has been seen that "fiducial probabilities" defined as above are, like "posterior probabilities" (determined by use of the "principle of insufficient reason"), cases of mathematical probability distributions, defined on the range of an unknown parameter i. The only substantive interpretation which the present writer can suggest for the term "fiducial probability" is that the term seems to be an instance of the tradition of broad usage of "probability", initiated by Bayes and Laplace in the <u>different</u> form discussed in the preceding Section 6, and used to express statements of informative inference about unknown parameters. It seems to the present writer that the problem of informative inference itself, for which evidently fiducial methods have been developed, is clarified by the analysis of Section 4 above, and served well and clearly by the intrinsic confidence methods defined and interpreted as in Section 5 above.

The scope of formal correspondence between intrinsic confidence methods and fiducial methods will be discussed for a wider class of problems in a following part of this paper. In the light of the preceding discussion, it will not be altogether unexpected if intrinsic confidence limits for the difference of means in the Behrens-Fisher problem exist and coincide in form with the fiducial limits given by Fisher for that problem.

8. The relativity of intrinsic evidential interpretations expressed in terms of error-probabilities.

Our use of simple cyclic-symmetric experiments as a frame of reference has played a technical role in establishing the principal conclusion of Section 4 above, concerning the basic status and role of the likelihood function for purposes of informative inference. In addition we have found, in Section 5, that simple c.s. experiments provide a convenient useful frame of reference for techniques, such as intrinsic confidence methods, which express some of the evidential meaning of likelihood functions. The intrinsic confidence coefficients associated with such intrinsic confidence statements, and analogous error-probabilities which may be associated with other such intrinsic evidential interpretations of experimental outcomes, are defined and meaningful only in association with the simple c.s. experiment (determined by the likelihood function) which may conveniently be adopted as an appropriate frame of reference for evidential interpretations.

Apart from convenience and simplicity, however, there is no reason of principle which recommends such frames of reference as uniquely appropriate for interpreting and expressing the evidential meaning of an observed likelihood function. The latter is basic and is itself evidentially meaningful, and its evidential meaning can be recognized in, and expressed in terms of, various alternative adequate experimental frames of reference.

For example, if $k = 2$, an outcome j which gives the likelihood function $(p_{1j}, p_{2j}) = (c, 99c)$, for any positive c, gives the likelihood ratio statistic the value 99, and can be interpreted

by use of the simple c.s. experiment (simple symmetric binary experiment) $E = \begin{bmatrix} .99 & .01 \\ .01 & .99 \end{bmatrix}$ as a conveniently chosen frame of reference. In the latter frame of reference, the outcome can be characterized as supporting H_2 against H_1 with an evidential strength associated with error-probabilities equal to .01. (Such an inference statement is formally an example of an intrinsic confidence method: On the basis of the outcome described, regardless of the structure of the experiment from which it was obtained, the hypothesis H_2 (or the set of parameter points i consisting of the single point i = 2) constitutes an intrinsic confidence set (or in this case an intrinsic confidence point) estimate, having intrinsic confidence coefficient .99).

However, the same outcome can be characterized just as properly, although perhaps less conveniently for some purposes, as evidentially equivalent to the second outcome of the asymmetric simple binary experiment $E' = \begin{bmatrix} 98/99 & 1/99 \\ 0 & 1 \end{bmatrix}$ in which "false negatives" are impossible but "false positives" have probability $1/99 \doteq .0101 > .01$.

The structure of E' is characterized by the two error-probabilities .0101, 0, while the structure or E is characterized by the two error-probabilities .01, .01. This example illustrates that when part of the evidential meaning of a likelihood function is expressed by use of intrinsically-associated error-probabilities (as in intrinsic confidence methods), the specification of the chosen experimental frame of reference (e.g. a simple c.s. experiment) must be included as an essential part of the interpretive

statements. No doubt, the choice of simple c.s. experiments, and their analogues in more general problems, will usually be convenient; and terms such as "intrinsic confidence coefficients" can be defined, as above, to refer automatically to such convenient frames of reference.

Part B: Translation and scale parameters.

9. Conditional inference methods. Let E denote any experiment having the following structure: Y is a random variable (r.v.) with c.d.f. $G(y - \theta)$, where G is known, and the unknown translation parameter lies in any specified subset Ω of the real line. (Alternatively, let Y^* be a positive random variable with c.d.f. $G^*(y/c)$, with G^* known and the scale parameter c unknown, $0 < c < \infty$. Then $Y = \log Y^*$ has c.d.f. $G(y - \theta)$ with translation parameter $\theta = \log c$, where $G(u) = G^*(\exp(u))$. Let $G(u) = \int_{-\infty}^{u} g(u)\,du$, $-\infty < u < \infty$. Let $x = (y_1, \ldots y_n)$ denote a sample of n independent observations on Y. Let $w = w(x) = (y_2 - y_1, \ldots, y_n - y_1)$; let $z = z(x) = y_1$; then (w,z) is a sufficient statistic, having a probability density function $h(w,z;\theta) = q(w)\; t(z - \theta;w)$, where the marginal density function $q(\cdot)$ of W, and the conditional density function $t(\cdot;\cdot)$ of $Z - \theta$, given that $W = w$, are known and independent of θ.

For each fixed w, let E^w denote the experiment consisting of a single observation z on the r.v. Z with p.d.f. $t(z - \theta;w)$ defined above, with unknown translation parameter θ. Let E_q denote the mixture experiment in which an observation w is taken on the r.v. W with p.d.f. $q(w)$, defined above, and then the experiment E^w is performed. A sufficient statistic for E_q is (w,z), which has p.d.f. $h(w,z;\theta) = q(w)\; t(z - \theta;w)$, as in E above. Thus $E = E_q$; that is, the experiment E is equivalent for all inference purposes

to the mixture experiment E_q. It follows that for typical purposes of informative inferences about θ, an outcome x of E can and should be interpreted as a corresponding outcome consisting of a single observation z from the experiment E^w, which has the known (conditional) p.d.f. $t(z - \theta;w)$.

Instead of a fixed number n of observations y_i as above, consider any sequential sampling rule, defined with reference only to the observed sequences of differences $w_2 = (y_2 - y_1), \ldots,$ $w_m = (y_2 - y_1, \ldots, y_m - y_1), \ldots,$ which terminates with probability one. For any sequence of observations $x = (y_1, y_2, \ldots)$, let $n = n(x)$ denote the number of observations y_i required for termination; then $n((y_1 - \theta), n(y_2 - \theta), \ldots))$ is a function independent of θ, which could be written $n(x) = n(w_2(x), w_3(x), \ldots)$. For each sequence x, let $z = y_1$ and let $w = w_{n(x)}$. Then as above we have that (w,z) is a sufficient statistic, with the distribution of w independent of θ, and the conditional p.d.f. of z having the form $t(z - \theta;w)$ with translation parameter θ. Thus the discussion and conclusion of the preceding paragraph is applicable also to such sequential experiments. One useful class of sequential sampling rules have the following form: Continue sampling until the form of $t(\cdot;w)$, which depends only upon w, allows (conditional) inferences about θ which are suitably highly inforative; e.g., until $t(\cdot;w)$ represents a translation-parameter family of distributions each of which is sufficiently highly concentrated to provide (conditional) confidence intervals for θ which are suitably short and have suitably high confidence coefficients, as determined in the following section.

10. <u>Intrinsic confidence methods</u>. If the range of θ is the real line, then the conditional frame of reference E^w described above is

an experiment characterized completely by the likelihood function $t(z - \theta; w)$ of θ on the observed outcome (w,z) of E; and the possible distributions of Z in E^w are a full translation group of distributions $(-\infty < \theta < \infty)$. Any conditional confidence methods of inference, in such a case, may also be called intrinsic confidence methods, in a natural extension of the usage introduced in Section 5 above.

$$\text{Let } F(u;w) = \int_{-\infty}^{u} t(u;w)\,du \text{ for each } u \text{ and } w.$$

Let $u(\alpha,w)$ be defined as the solution u of the equation $\alpha = F(u,w)$, and let $\theta(z,w,\alpha) = z - u(\alpha,w)$, for each pair α,w for which the first equation has a unique solution. Then $\theta(z,w,\alpha)$ is a lower α-level confidence limit estimator (and/or an upper $(1 - \alpha)$-level upper confidence limit estimator) of θ (conditional on w). $\theta(z,w,.5)$ is a median-unbiased point-estimator of θ. The pair $\theta(z,w,.95)$, $\theta(z,w,.05)$ is a 90 o/o confidence interval estimator of θ.

For each constant $k \geq 0$ and each w, let

$$A(w,k) = \{u \,|\, t(u,w) > k\}, \text{ and let } \gamma(w,k) = \int_{A(w,k)} t(u;w)\,du.$$

For each z, w, and k, let $B = B(z,w,k) = \{\theta \,|\, (z - \theta) \in A(w,k)\}$. Then B is a γ-level confidence set estimator of θ (conditional on w). Such estimators may be called maximum likelihood (conditional) confidence sets; it is easily verified that for each γ, such set estimators are optimum in the sense that they have (conditionally and unconditionally) minimum Lebesgue measure, among all set estimators whose confidence coefficients, conditional on w, are never smaller than γ.

11. <u>Discussion</u>. It is interesting that such conditional confidence limits and sets coincide in all formal details (though not in interpretation) with those based upon the traditional Bayesian "principle of insufficient reason" in which, after initial reference to a "uniform prior distribution of Θ over $-\infty < \Theta < \infty$," the likelihood function $t(z - \Theta;w)$ is treated formally as a posterior p.d.f. of Θ.

Furthermore, within the conditional frame of reference of an experiment E^w, which seems appropriate for purposes of informative inference, z is a sufficient statistic, and the quasistatistic $v(z,\Theta) = z - \Theta$ has the same distribution under each Θ. Hence it is evidently possible to apply formally the "fiducial argument" used by Fisher to define a "fiducial probability distribution of Θ." It is interesting that the resulting "fiducial p.d.f. of Θ" coincides with the likelihood function $t(z - \Theta;w)$, and inference statements about Θ of the fiducial type coincide in all formal details (though not in interpretation) with those obtained above as conditional confidence statements about Θ.

The principal conclusion of the preceding sections may be stated briefly as follows: For purposes of informative inference concerning a translation parameter Θ, regardless of the structure of an experiment E (so long as Θ is a translation parameter with respect to the distributions of outcomes of E), the appropriate frame of reference is determined just by the likelihood function on the outcome observed: this frame of reference is an experiment (generally different from E) in which all possible outcomes give likelihood functions differing from that observed only by a translation. We recall that, as in Sections 5 and 8 above, each

conditional (or intrinsic) confidence method gives only a partial summary and interpretation of the likelihood function, and that the latter, with the totality of its possible interpretations, is basic to informative inference. (A logarithmic transformation reduced scale-parameter problems to translation-parameter problems.)

The usual general approach to point-and confidence-interval estimation, for example as formulated and developed in [2], takes the given experiment E as the basic frame of reference for inference methods and statements. Examples 1, 2, 4, 5, 7, and 8 of [2] involve translation or scale parameters, and in all but the first two examples the confidence methods given there differ markedly from those developed above. For purposes of informative inference, the formulation and methods of the present paper seem preferable in principle, for the reasons given above.

The methods of the preceding sections for translation and scale parameters have important points of agreement and of difference with those given by Pitman [3].

The methods of the preceding sections for translation parameters admit immediate generalization to multiparameter problems, such as a p.d.f. $g(y_1-\theta, y_2-\theta)$ of a random vector $Y = (Y_1, Y_2)$. Analogous methods apply to rotation parameters such as the case $Y = (R,Z)$, with p.d.f. $g(r, z-\theta)$, $0 \leqq r < \infty$, $0 \leqq z-\theta < 2\pi$, where (r,z) are the polar coordinates of a point y in the plane. A specific example is the case of a bivariate normal distribution, with identity covariance matrix and with mean lying on the unit circle.

12. Acknowledgements.

The writer is indebted to Prof. D.A.S. Fraser for the observation that a version of the "fiducial argument" is applicable to any simple symmetric binary experiment; and to Prof. G.A. Barnard and Mr. J. Cornfield for a number of helpful suggestions and discussions during the preparation of this paper. (This is not to suggest that the general views of these colleagues are necessarily represented in this paper.)

REFERENCES

[1] Allan Birnbaum, "On the foundations of statistical inference. I. Binary experiments," Technical Report IMM-NYU 267, 1960, Institute of Mathematical Sciences, New York University. To be published.

[2] Allan Birnbaum, "A unified theory of estimation. I.," Technical Report IMM-NYU 266, 1960, Institute of Mathematical Sciences, New York University. To be published (slightly abridged) in the Annals of Mathematical Statistics, Vol. 32 (1961).

[3] E. J. G. Pitman, "The estimation of the location and scale parameters of a continuous population of any given form," Biometrika, Vol. 30 (1938), pp. 391-421.